Why does caste still influence access to agricultural credit?

Sunil Mitra Kumar*

May 2016

Abstract

In India, caste shapes access to a variety of resources and outcomes mainly through its influence on inter-generational prosperity, but also the linked phenomenon of discrimination. This paper examines whether caste-based differences in access to formal agricultural loans reflect discrimination in banks' lending. We find, first, that there are significant caste-based differences in loan application rates. Having controlled for the decision to seek credit and various borrower characteristics, we also find that loan approval rates across castes are largely equal across three of four caste-groups, but that Scheduled Tribe (ST) borrowers are less likely to have loans approved. Through a simulation-based approach, we show that these lower approval rates for STs are most likely not explained by unobserved credit histories, suggesting that banks do discriminate against STs. We discuss the policy implications that arise from these findings.

*India Institute, King's College London, Strand, London WC2R 2LS. E-mail: Sunil.Kumar@kcl.ac.uk.
I am grateful to Shaun Hargreaves-Heap for comments, and to the United National University World Institute for Development Economics Research for financial support. The usual disclaimer applies.

1 Introduction

Caste as a form of social identity remains an enduring predictor of economic status in India. It is correlated with occupation and employment (Prakash, 2015; Ito, 2009; Thorat and Attewell, 2007), income and expenditure (Deshpande, 2000), and capital more generally (Kijima, 2006). Caste thus remains an important indicator of socioeconomic disadvantage. One important avenue through which caste can influence income and productivity in the rural sector is by shaping access to agricultural credit. Agriculture is the largest employer in India, and due to the gap between sowing and harvest credit is usually needed to be able to purchase various inputs (Conning and Udry, 2007). The question of whether caste influences access to credit is thus important, and while there is a general agreement that it does (Kumar, 2013; Pal, 2002; Government of India, 2007), the reasons for these differences are less clear. One particular concern is whether lenders discriminate on the basis of caste (Kumar, 2013; Dréze et al., 1997).

This paper delves into why caste influences access to agricultural credit, and whether banks discriminate on the basis on caste. We use nationally-representative data from the 2011-12 round of the India Human Development Survey (Desai and Vanneman, 2015), which enables us to study the decision to apply for a loan separately from the subsequent approval of this application. Previous studies in this area have studied the determinants of access to credit including caste (Kumar, 2013; Pal, 2002) and whether farmers are credit-constrained (Kochar, 1997). These studies control for credit-worthiness characteristics including land-ownership, household demographics, and indicators of economic status, but do not control for the decision to seek credit. To our knowledge, our paper is the first attempt to study agricultural lending using data that include information about loan applications, thus enabling us to explicitly distinguish between lender-based differences in loan approval rates and borrower-based differences in the demand for credit.

To make sense of caste-based discrimination in accessing credit, it is useful to note the parallels with discrimination in labor markets (Thorat and Attewell, 2007; Altonji and Blank, 1999) and the role of racial and gender-based discrimination in accessing bank credit in the USA (e.g. Blanchflower et al., 2003; Blanchard et al., 2008; Asiedu et al., 2010). The definition of discrimination commonly employed in economics to study this phenomenon was proposed by Gary Becker (1971). Discrimination occurs if the differences in some outcome (loan approvals, wages) based on race, caste, or a similar attribute, are not 'objective'. That is, if residual differences remain even after taking into account all possible borrower or worker characteristics that are relevant to the outcome, and are attributed to a 'taste' for discrimination on the part of the bank or employer. Vice-versa, discrimination due to 'objective' reasons – credit worthiness, ability – can be called statistical discrimination. A third way in which outcomes might differ systematically by caste or gender or race is if expectations depend on belonging to a certain such category. For instance, members of one group might be less likely to expect, and thus demand, better pay (Chevalier, 2007).

In this paper we attempt to distinguish between these three phenomena. Castes are usually categorised into four major groups, viz. Scheduled Castes (SCs; the most disadvantaged), Other Backward Classes(OBCs; of middling disadvantage), Brahmins/Others, the traditionally privileged higher castes, and Scheduled Tribes (STs; also very disadvantaged).[1,2,3] Analysing the decision to apply for credit, we are able to estimate caste-wise residual differences in loan application rates having controlled for various indicators of socio-economic status.

Our first main finding is that a major part of caste-wise differences in access to loans can be explained by differences in application rates. These differences mirror patterns of socioeconomic disadvantage, but persist even after accounting for socioeconomic status.

[1]There is substantial and enduring complexities and political agitation around the categorisation of castes as OBC (Ramaiah, 1992).

[2]STs are not technically part of the caste system. Yet tribals are one of the most disadvantaged peoples in India, and most analyses of caste therefore include the ST category, as does the current paper.

[3]It is generally agreed that the caste system applies not only to Hindus but to followers of other religions in the Indian subcontinent as well.

SCs and STs, the two most disadvantaged groups, are 16-20% less likely to apply for loans than are Brahmins and OBCs. We cannot explain *why* these differences exist, but a likely explanation seems that they reflect expectations about loan approval, which therefore reflect caste-wise disadvantage in other realms of life.

Next, by analysing loan approvals for those who applied, we aim to isolate taste-based discrimination from plausible sources of statistical discrimination. As a first attempt, we control for the covariates plausibly linked to credit-worthiness in an agricultural context, and interpret caste-wise residuals as evidence of lenders' taste discrimination. Our second main finding is that once the decision to seek credit is controlled for, Brahmins, OBCs and SCs all have near-equal chances of having a loan approved, but that ST borrowers are 5-7% less likely to have a loan approved.

Yet our data are observational, and it is possible that an important unobservable might lead to an underlying caste-wise statistical difference being (mis)interpreted as evidence of taste-based discrimination. We examine this possibility through a simulation approach, and we believe this is an important methodological contribution of the paper. We adapt an approach suggested by Ichino et al. (2008) and Rosenbaum and Rubin (1983) to simulate a binary 'credit history' variable, since credit histories are not recorded in our data, but could be important determinants of loan approvals. We then study how the residual effect of caste-group belonging on loan approvals changes with different distributional assumptions for these simulated credit histories. Through this exercise, we argue that the lower loan approval rates for STs are likely robust to STs being poorer credit risks as a group, or in other words, that STs do likely face taste-based discrimination from banks.

Finally, we divide our data into subpopulations and analyse these separately to further investigate caste-based differences. We find that the apparent discrimination against STs is visible mainly in states with substantial ST populations, and not where they are a small minority. We also divide the sample according to land ownership to indirectly examine affirmative action guidance provided by the Reserve Bank of India (RBI).While caste-based reservations in government jobs and in admissions to educational institutions (Bertrand et al., 2010) are more prominent affirmative action initiatives, the RBI has long directed bank lending towards disadvantaged groups (Reserve Bank of India, 2014, 2004, 2008; Sriram, 2007). These policies are especially significant for agricultural lending in rural India, where expanding formal bank credit is thought to be key to increasing rural incomes, and they encourage banks to lend to 'small' farmers, viz. those who own at most 5 acres of land. We find indirect evidence as to the success of this guidance, since caste-wise patterns for loan applications and approvals are very muted for small farmers, and the overall patterns of difference are in effect explained by lending patterns amongst larger farmers.

The reminder of the paper is structured as follows. Section 2 presents the data, while section 3 lays out our empirical approach. Results are presented in section 4, and section 5 discusses their implications.

2 Data

We use data from the second round of the India Human Development Survey, a nationally-representative household survey undertaken in 2011-12 (Desai and Vanneman, 2015). These data provide information on a rich set of social, economic and demographic characteristics which are well suited to analysing the determinants of households' borrowing and debt from formal and informal sources. We focus only on bank, i.e. formal sector loans in the current paper. Such loans are widely considered superior than what the informal sector provides, owing to the usurious lending practices in the latter. The question of caste-based differences is therefore of policy and more general importance for the formal sector. It is also usually assumed that given a choice, farmers would opt for formal sector loans (e.g. Kochar, 1997). Therefore, excluding informal sector borrowings is also unlikely to compromise the identification of caste-based differences in application and approval rates for formal sector loans.

This survey is especially useful since households were asked not only to list what loans they did have, but crucially, whether they had applied for loans from various sources, and the success of those applications. In the Indian context this question is critical to exploring whether any caste-wise differences in loan patterns do indeed reflect discrimination on the part of lenders, and to our knowledge, no other large-scale household survey in India has asked this question.[4] As per the questionnaire, details about agricultural bank loans in terms of outstanding loans, applications, and approvals, refer to a loan taken during the five years preceding the survey from a bank for the stated purpose of furthering agricultural business or purchasing agricultural equipment.

The second round of the IHDS data currently consist of individual and household-level information, but village-level information are covered only in the first round of the survey that took place in 2004-05. In particular, the presence of a bank in the village or distance to the nearest such is potentially an important covariate that helps control for loan supply. But, since the data on village covariates are from the earlier survey round, we present results with and without including them.

We focus on farmer households, which we define as those households who report that a) their main occupation is cultivation, and b) that they cultivate land. In terms of regions, we exclude Jammu and Kashmir, the North-Eastern states, and small states and Union Territories since these account for very few loans. Table 1 presents means and proportions for all variables by caste-group, and all statistics shown are population estimates that take survey weights into account, as do all our estimations.

Caste-wise means and proportions follow expected patterns in that land ownership, rent-

[4]The National Sample Survey Organisation's All India Debt and Investment Survey is the other main source of information on household's access to credit. This is a decennial survey, conducted most recently in 2013. In keeping with previous rounds, it asks about the details of existing loans but does not ask about loan applications.

Table 1: Means and proportions by caste-group

N=7,510	Brahmin	OBC	SC	ST
Loans				
Proportion currently have loan	0.326	0.337	0.215	0.169
Proportion applied for loan	0.353	0.362	0.238	0.207
Proportion loan approved if applied (N=2,587)	0.922	0.930	0.901	0.815
Land and income				
Land owned (acres)	15.824	11.819	7.028	10.187
Land rented in (acres)	2.156	2.633	2.303	1.573
Land rented out (acres)	1.912	1.539	1.262	1.186
Monthly consumption expenditure (Rs. '000)	128.383	105.529	82.755	61.398
Annual income (Rs. '000)	129.444	88.313	64.671	50.677
Education and household composition				
Age of household head	53.533	52.019	51.067	49.269
Proportion with male household head	0.943	0.945	0.945	0.916
Years of education highest male	8.529	7.213	6.108	5.399
Years of education highest female	5.971	4.132	3.462	2.680
Household size	5.298	5.360	5.216	5.258
Household proportion of adult males	0.379	0.353	0.340	0.340
Household proportion of adult females	0.365	0.354	0.348	0.344
Village characteristics				
Distance to nearest bank branch (km)	5.784	5.928	6.590	8.573
Distance to nearest town/city (km)	14.328	15.559	16.538	22.136
Percentage of households with electricity	63.317	58.907	55.715	45.742
State population proportions				
Himachal Pradesh	0.667	0.084	0.222	0.027
Punjab	0.730	0.169	0.101	0.000
Uttarakhand	0.228	0.234	0.538	0.000
Haryana	0.559	0.362	0.076	0.003
Rajasthan	0.181	0.545	0.156	0.118
Uttar Pradesh	0.279	0.593	0.127	0.000
Bihar	0.275	0.620	0.091	0.014
West Bengal	0.483	0.184	0.297	0.036
Jharkhand	0.134	0.266	0.100	0.499
Orissa	0.099	0.511	0.174	0.215
Chhattisgarh	0.011	0.481	0.078	0.430
Madhya Pradesh	0.229	0.477	0.094	0.201
Gujarat	0.327	0.441	0.049	0.183
Maharashtra	0.362	0.475	0.062	0.101
Andhra Pradesh	0.311	0.485	0.181	0.023
Karnataka	0.284	0.565	0.070	0.081
Kerala	0.637	0.303	0.061	0.000
Tamil Nadu	0.069	0.720	0.191	0.021

Notes:
This table presents means and proportions by caste group. All statistics are estimates for the population that take into account survey probability weights.

ing and leasing, incomes, and access to credit are all higher for Brahmins and OBCs, followed by SCs and then STs. It notable that the figure for incomes is *lower* than that for consumption expenditure, even though the former is an annual figure and the latter monthly. The survey documentation (Desai and Vanneman, 2015) offers two reasons for this: first, several incomes are negative, reflecting the precarious nature of cultivator livelihoods. Second, calculating incomes is notoriously difficult when several components are measured in kind.[5]

3 Empirical framework

Our analysis seeks to estimate the residual effects of caste belonging on the demand for loans, as well as their subsequent approval. We use Logit specifications to model both the decision to apply for an agricultural loan, and conditional on this, the bank approving this application.

For household i, let Y_{1i} and Y_{2i} denote the binary outcomes of applying for an agricultural bank loan and, respectively, having this application approved and thus receiving a loan. We model the probability of each outcome using a logit specification:

$$\text{Prob}[Y_j = 1|\mathbf{X}] = \frac{e^{\mathbf{X}'\beta}}{1 + e^{\mathbf{X}'\beta}} \qquad \text{where} \quad j \in \{1, 2\} \tag{1}$$

Here $\mathbf{X}$ is a vector of covariates that includes caste-group dummies and β is a vector of coefficients that we seek to estimate.

Unbiased estimation of the residual effects of caste-group on loan applications and the approval thereof requires controlling for household characteristics that influence the decision to apply for credit, and, respectively, to the bank's decision to approve that application.

[5]This pattern is also borne out by data from the National Sample Survey Organisation. See table 2 in Mishra (2008)

To this end we include variables known to influence access to credit (see, for example, Pal, 2002; Swain, 2007), including household income and land owned – the latter the primary form of collateral demanded by banks.[6]

Second, we must control for the supply of loans. For loan applications, distance to the nearest bank (including if such is present in the village) is an important indicator both of supply as well as a potential proxy of a bank's knowledge about potential borrowers. We include this distance as a village-level covariate, together with two indicators of development and connectedness, viz. the proportion of households with electricity, and the distance to the nearest town/city. Unfortunately the data on these covariates are not ideal, since they are from an earlier round of the IHDS survey conducted in 2004-05.

Analyses of racial discrimination in access to mortgage or small business loans typically also control for past repayment histories and default behaviour (e.g. Blanchflower et al., 2003). Since credit ratings and the bureaus that maintain such information are not prevalent in the rural Indian context, it is unlikely that banks use such information systematically. Nonetheless, credit histories or ratings are a potentially important influence on loan applications, and to the extent that they might be correlated with caste-group, their exclusion could lead to biased estimates. In particular, information on repayment rates or credit histories is important to distinguish between statistical discrimination (owing to caste-repayment correlation), and taste-based discrimination. We can infer the latter should residual caste-wise differences persist even after controlling for repayment histories.

However, the data do not provide information on repayment histories or a suitable proxy. Instead, we therefore examine their potential role through a Monte-Carlo simulation technique adapted from Ichino et al. (2008). This approach tests the sensitivity of our estimation results to the presence of an unobservable. Section 4.1 details the method,

[6]According to the Reserve Bank of India's guidelines, banks should not demand collateral for loans of up to Rs 50,000, and for larger loans, land is the main form of collateral (Reserve Bank of India, 2007).

and the intuition is as follows. Excluding credit histories would lead to biased estimation provided credit histories are correlated with caste-group as well as a bank's decision to lend. Thus, we can simulate a variable that is correlated with both a caste-group of interest and loan approvals and re-estimate our results. Using Monte-Carlo draws and varying the correlation between the simulated variable and, respectively, caste-group and loan approval, we can study the distribution of caste-group residuals by repeatedly re-estimating the model. The difference between these distributions and the estimates obtained without the simulated variable provide a handle on the sensitivity of results, and thus the potential importance of credit histories.

4 Results

The first three rows of table 1 provide (unadjusted) caste-wise proportions in terms of access to credit. The first row confirms that there exist substantial inter-caste differences in loan access, and in particular, SC and ST borrowers have relatively low access to credit. These proportions are the type of statistic provided in official reports such as National Sample Survey Organisation (2005), and previous analyses essentially examine how these proportions change once borrower characteristics are accounted for, but do not differentiate between loan applications and their approval (e.g. Kumar, 2013). In contrast, this paper aims to investigate the role of caste in both applications and approvals. The second row shows that inter-caste differences in loan applications might indeed explain patterns of overall access in part. Third row suggests that loan approvals are also not uniform across all caste-groups: while around 90% of loan applications by Brahmins, OBCs and SCs are approved, the proportion for STs is only 81.5%. Our empirical approach thus aims to rigorously examine the loan application and approval statistics, and analyse the extent of residual inter-caste differences that remain once we control for household characteristics that proxy ability, credit-worthiness, and access.

Table 2: Logit estimation results

Model	Loan applications			Loan approvals		
	OBC	**SC**	**ST**	**OBC**	**SC**	**ST**
(1) State dummies	1.021	0.555	0.467	1.099	0.824	0.369
	(0.107)	(0.074)	(0.071)	(0.304)	(0.294)	(0.136)
		p=0.000			p=0.009	
(2) Model 1 plus land and income[a]	1.222	0.802	0.656	1.252	1.033	0.454
	(0.126)	(0.111)	(0.106)	(0.333)	(0.360)	(0.163)
		p=0.000			p=0.026	
(3) Model 2 plus education and [b] household characteristics	1.260	0.846	0.722	1.387	1.299	0.541
	(0.131)	(0.119)	(0.120)	(0.359)	(0.427)	(0.197)
		p=0.000			p=0.057	
(4) Model 3 plus village characteristics[c]	1.255	0.837	0.746	1.386	1.293	0.542
	(0.130)	(0.118)	(0.130)	(0.366)	(0.427)	(0.186)
		p=0.000			p=0.033	

Notes:
This table presents estimates from logit models. Only caste dummy coefficients are shown, with the full regression results shown in appendix A. Coefficients are given in terms of odds ratios, with standard errors in parentheses. The base category is Brahmins. All estimates are survey-weighted, and standard errors in parentheses account for village-level clustering as per the survey design. The p-values shown are for Wald tests for the null that all three caste dummies are jointly zero. Sample size is 7,510 for loan application models and 2,587 for loan approval models.

[a] Land owned, land rented out, land rented in, household monthly consumption expenditure, and annual income.

[b] Age and sex of household head, years of education of highest-educated male and female, household size, household proportion of adult males and females.

[c] Distance to nearest bank branch, distance to nearest town/city, proportion households with electricity.

Table 2 presents the estimation results from a series of Logit models. These are in terms of odds ratios for respective caste dummies (full regression results are presented in appendix A), and the bottom part of each row reports p-values from Wald tests of joint equality of all three caste dummies.

We estimate two sets of models, one for loan applications (columns 2-4) and one for their subsequent approval by a bank resulting in a loan (columns 5-7). In all cases the Brahmin caste group is the base category. For instance, the first row says that in a simplistic model that includes only state dummies, the odds ratio for an SC household applying for a loan (respectively, having that application approved) relative to Brahmins is 0.555 (respectively, 0.824). This simple model thus mirrors the unadjusted caste-wise proportions in table 1.

Subsequent rows in table 2 show how respective odds ratios change as additional variables are added to the model. Row 2 adds household-level characteristics including land owned, rented out and rented in, and income and consumption expenditure, while row 3 adds the age and sex of the household head, the years of education of the most educated male and female members, household size, and the proportion of adult males and females. Row 4 adds village-level information on distance to the nearest bank and to the nearest town/city, and the proportion of households with electricity. In general, as we move from simpler to more complex models, inter-caste differences in both applications and approvals are reduced as more covariates are controlled for. The odds ratios become closer to unity even though, as the p-values suggest, substantive inter-caste differences persist.

To examine how these estimation results relate to unadjusted the caste-wise proportions in table 1, we calculate the predicted probabilities of loan application and approval. We report the sample-averaged predicted probabilities corresponding to model 4 in table 3 (those for model 3 are very similar, as the odds ratios would suggest). For these

Table 3: Predicted probabilities by caste group

Loan applications				Loan approvals			
Brahmin	OBC	SC	ST	Brahmin	OBC	SC	ST
0.315	0.356	0.285	0.266	0.908	0.930	0.926	0.852
(0.015)	(0.012)	(0.020)	(0.023)	(0.017)	(0.009)	(0.015)	(0.033)
p=0.000				p=0.033			

Notes:
This table shows sample-average probabilities by caste group predicted by model 4 in table 2.
Standard errors are given in parentheses, and p-values shown are for Wald tests for the null that the predicted probabilities are equal across all four caste categories.

predictions, all variables except caste are held at their sample values, and the caste variable is changed to Brahmin (say) to calculate the predicted probabilities for Brahmins, and similarly for other caste-groups. The decision to hold all non-caste variables at their sample values (rather then means, for instance) ensures that these calculations have a genuine interpretation as the actual probabilities that would have resulted were all households Brahmin (and so on). The bottom half of each row provides results from Wald tests for comparisons of these caste-wise predictions, and these show that inter-caste differences are also statistically significant.

Thus, columns 2-5 of table 3 show that loan applications vary by caste, in a way that mirrors unadjusted proportions (table 1) but with attenuation. This pattern confirms that residual inter-caste differences in loan application rates are an important factor behind differences in overall access to credit. It also suggests, as we discuss in section 5 in detail, that attempts to reduce inter-caste disparities in this access are unlikely to yield swift results, since at least for SC and ST borrowers, these inter-caste differences might well reflect historical experiences of disadvantage and discrimination.

Next, columns 6-9 show that Brahmins, OBCs and SCs all have similar rates of loan approval, but that the corresponding rate for STs is 5-7% lower. This marked residual difference despite controlling for a rich set of relevant covariates suggests that banks might discriminate against STs. However, while we have controlled for important components of credit-worthiness – in particular in the form of collateral (land) and incomes – the data do not have information credit histories. We now investigate this issue through a simulation approach, to gauge the potential influence of unobserved credit histories on caste-group residuals for loan approvals.

4.1 A simulation approach to unobserved credit histories

Suppose that credit histories influence banks' decisions to approve agricultural loans, and that as a group, STs have inferior credit histories compared to non-STs. Then, the residual negative effect of ST group-belonging on loan approvals might in fact get partially or fully explained away by controlling for loan histories in equation 1. In other words, while our results suggest that banks discriminate against STs, the absence of this variable in the data means that we cannot infer whether the discrimination is statistical – if STs have worse credit histories on average – or taste-based. This is a crucial question from the perspective of understanding how banks, a central financial institution in rural India, approach the question of caste, and indeed whether there are parallels here with race and gender-based discrimination in other contexts.

We approach this problem using Monte-Carlo simulation, adapting an approach suggested by Ichino et al. (2008) and Rosenbaum and Rubin (1983). Both papers investigate how including a hitherto binary unobservable might change the estimate of a treatment effect. They assume that assignment to treatment is not unconfounded if the unobservable is excluded, but becomes so once it is controlled for. Ichino et al. (2008) implement this idea for treatment effects estimation using propensity score matching, and propose a simple way of parameterising the distribution of the binary unobservable. Partitioning the set of observations into four quadrants according to (binary) treatment status and (also binary) outcome, four corresponding probabilities that the unobservable equals one are specified.

Monte-Carlo draws are used to construct the unobservable such that it conforms, on average, to this set of probabilities. The treatment effect is then re-estimated for each draw having included the unobservable as a covariate, and this exercise is repeated across multiple draws and with different combinations of the probabilities characterising the unobservable.

We adapt Ichino et al. (2008)'s method for characterising the distribution of the unobservable, even as our set-up is more similar toRosenbaum and Rubin (1983) who use a logit model to estimate the effect of a binary treatment on a binary outcome. In our case, 'treatment' is the categorical variable of caste-group belonging, and outcome is the binary loan-approval. Caste-group cannot be manipulated in reality, but conceptually we are studying the consequences that would arise as if it could. And any caste-wise differences in loan approval calculated by averaging over the full sample (as we did in table 3) are in effect estimates of this treatment effect.

Further, our focus is on potential discrimination against STs. Therefore, while we include all four caste categories when analysing caste-wise probabilities of loan approval, we simplify caste to STs and non-STs for the purpose of constructing the unobservable 'credit histories'. This maintains tractability, but more importantly, is intuitively appealing: our results thus far suggest that differences in loan approvals between Brahmins, OBCs and SCs are negligible, and it is the gap between these groups and STs that potentially suggest institutional discrimination.

4.1.1 Constructing the unobservable

Let H denote hitherto unobserved credit histories, where for household i, H_i can either be 1 (good history) or 0 (poor history). Let ST denote ST-group belonging with $ST_i = 1$ for STs, and $ST_i = 0$ all other caste-groups. And, let Y denote the loan outcome, with $Y_i = 1$ if the loan is approved and $Y_i = 0$ if it is denied.

We can now partition the set of observations as $\{Y = i, ST = j\}, i, j \in \{0, 1\}$. For instance, $\{Y = 0, ST = 1\}$ denotes the set of households who are ST and who have been denied loans. Following Ichino et al. (2008), let p_{ij} be defined as

$$p_{ij} = Prob(H = 1|Y = i, ST = j), i, j \in \{0, 1\} \tag{2}$$

Then the set $\{p_{00}, p_{01}, p_{10}, p_{11}\}$ fully characterises the distribution of H.

Now, let θ_{ij} denote the proportion of the sample in each of the sets $\{Y = i, ST = j\}$ where $i, j \in \{0, 1\}$. And let the proportion of STs (non-STs) with good credit histories be denoted μ_{ST} (respectively, μ_{nonST}).

Then

$$p_{11}\theta_{11} + p_{01}\theta_{01} = \mu_{ST} \tag{3}$$

$$p_{10}\theta_{10} + p_{00}\theta_{00} = \mu_{nonST} \tag{4}$$

An intuitive way of interpreting these numbers is as follows.

Suppose that banks base loan approval at least in part on the presence of a good credit history, and let $Prob(Y = 1|H = 1, ST = i) = b_{ST=i}$ where $i \in \{0, 1\}$. Taste-based discrimination on the part of banks would lead to different values of b for STs and non-STs, whereas purely statistical discrimination would not. To be conservative in our analysis in an attempt to see if statistical discrimination can be ruled out, we assume a single value $b = b_{ST=1} = b_{ST=0}$ for all households. Then, $p_{11} = p_{10} = b$.

Also, we want to investigate whether STs having poorer credit histories could be an explanation for the apparently lower rates of loan approval. Therefore we assume that $\mu_{nonST} > \mu_{ST}$. Let $ST_{gap} = \mu_{nonST} - \mu_{ST}$, i.e. the proportion by which non-STs have better credit histories than do STs.

We can now rewrite equations 3 and 4 as

$$b\theta_{11} + p_{01}\theta_{01} = \mu_{ST} \tag{5}$$

$$b\theta_{10} + p_{00}\theta_{00} = \mu_{ST} + ST_{gap} \tag{6}$$

Since p_{01} and p_{00} can be calculated using 5 and 6, the distribution of H can be fully characterised in terms of three parameters: the degree to which banks base loan approvals on credit histories (b), the proportion of STs with good credit histories (μ_{ST}), and the gap between this proportion and the corresponding (assumed higher) proportion for non-STs (ST_{gap}).

Note that not all combinations of $\{b, \mu_{ST}, ST_{gap}\}$ are feasible. Rearranging 5 and 6 we get

$$1 \geq p_{01} \geq 0 \quad \Leftrightarrow \quad 1 \geq \frac{\mu_{ST} - b\theta_{11}}{\theta_{01}} \geq 0 \tag{7}$$

$$1 \geq p_{00} \geq 0 \quad \Leftrightarrow \quad 1 \geq \frac{\mu_{ST} + ST_{gap} - b\theta_{10}}{\theta_{00}} \geq 0 \tag{8}$$

Furthermore, since each of b, μ_{ST}, and $\mu_{ST} + ST_{gap}$ are proportions:

$$1 \geq b \geq 0 \tag{9}$$

$$1 \geq \mu_{ST} \geq 0 \tag{10}$$

$$1 \geq \mu_{ST} + ST_{gap} \geq 0 \tag{11}$$

In general, ST_{gap} cannot be very high unless b is high too. The θ terms are all sample statistics. For STs, we have $\theta_{11} = 0.815$ and $\theta_{01} = 0.185$, while for non-STs we have $\theta_{10} = 0.925$ and $\theta_{00} = 0.075$. We restrict $\{b, \mu_{ST}, ST_{gap}\}$ to multiples of 0.01 between 0.01 and 0.99, and appendix B provides a graphical summary of all feasible $\{b, \mu_{ST}, ST_{gap}\}$ combinations where feasibility is defined by equations 7-11. The most 'adverse' feasible combination in terms of maximum ST_{gap} is $\{b, \mu_{ST}, ST_{gap}\} = \{0.98, 0.8, 0.18\}$.

4.1.2 Results

Based on the logic of omitted variables in standard regression analysis, if credit histories do explain residual inter-caste differences in loan approvals then once H in included in the model we would expect:

1. For a given b and μ_{ST}, the residual gap between STs and non-STs will decrease with higher values of ST_{gap}. This is because more of what used to be a residual gap would in fact be accounted for STs having worse credit histories.

2. For given b and ST_{gap}, the residual gap between STs and non-STs will decrease with lower values of μ_{ST}. Since the ratio of non-STs with good histories to that of STs is $\frac{\mu_{ST}+ST_{gap}}{\mu_{ST}}$, this ratio increases the smaller μ_{ST} gets. Again, the inclusion of H will account for more of the ST v. non-ST gap the larger this ratio, or the smaller μ_{ST}.

We reestimate our results for several specifications of H in terms of $\{b, \mu_{ST}, ST_{gap}\}$. For a given combination of $\{b, ST_{gap}\}$, we select the minimum feasible μ_{ST} so as to maximise the potential confounding influence of H. H is simulated using Monte Carlo draws from distribution corresponding to each $\{b, \mu_{ST}, ST_{gap}\}$ combination 300 times. For each draw, we then re-estimate the Logit model for loan approval having included H as an additional regressor. All other available covariates (model 4 in table 2) are also included.

For each regression, we calculate the sample-average caste-wise probabilities. The result in table 3 suggested that STs were 5-7% less likely to receive a loan than non-STs

(Brahmins, OBCs, SCs). To compare the results from the simulation exercise that includes H with this original result, we summarise the ST vs non-ST difference in loan approval probabilities using box plots. To also account for the standard errors of these predictions, we also conduct a Wald test each time for the null that caste-wise average predicted probabilities are equal.

Figure 1 summarises these results. The top box plot in each case summarises the ST vs non-ST gap in predicted loan approval probabilities, while the bottom box plot summarises the p-values for corresponding Wald tests. These show that the gap in predicted probabilities of loan approval between STs and non-STs remains robust to the inclusion of H for many but not all combinations of $\{b, ST_{gap}\}$. If loans are (very) strongly conditional on good credit histories ($b = 0.98$), then STs no longer have lower rates of loan approval with even modest differences between their credit histories and those of non-STs ($ST_{gap} = 0.18, 0.15, 0.1, 0.05$).

That said, $b = 0.98$ implies that good credit histories would be compulsory for loan approvals, and this level of dependence would appear implausible in the rural Indian context. For lower levels of b, i.e. the dependence between loan approvals and credit histories, the gap in loan approvals for ST and non-STs stays relatively unchanged at about 5%, and for some cases (e.g. $b = 0.4$) the range of simulated values includes higher gaps as well. Throughout, the median p-value for these comparisons is no more than 0.05 with three exceptions ($b = 0.8, ST_{gap} = 0.1; b = 0.6, ST_{gap} = 0.1; b = 0.4, ST_{gap} = 0.05$).

Appendix B presents corresponding box plots for the same estimations that summarise the gap between STs and the caste group with the minimum average predicted probability. This is a more conservative estimate of the ST vs non-ST gap, and these results confirm the pattern as figure 1, but with a gap that is about 1% less. The same appendix also presents boxplots of p-values from Wald tests for the joint equality of predicted probabilities across all four caste-groups. These are also very similar to the p-values in figure 1.

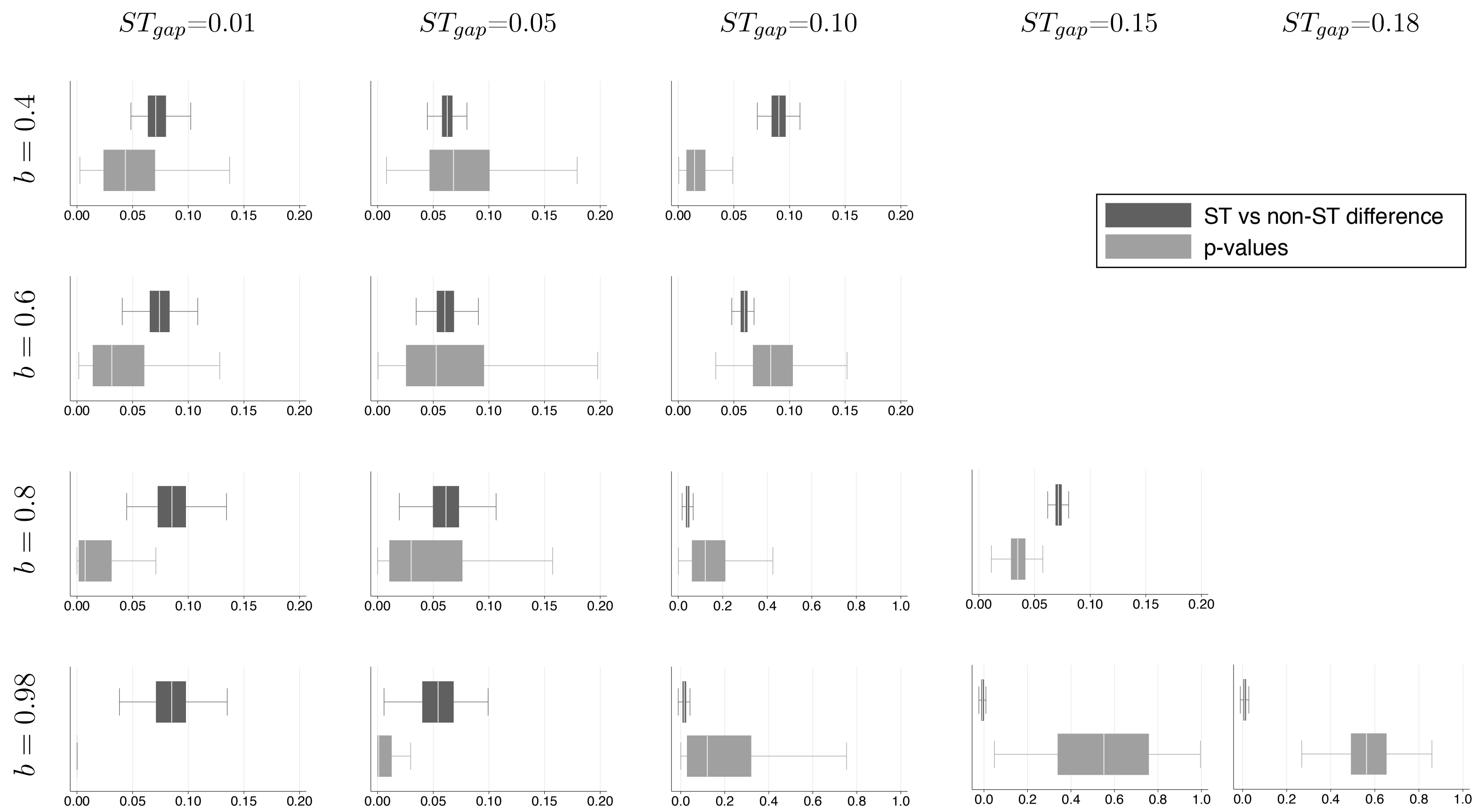
Do banks discriminate against ST borrowers? The results from this simulation exercise suggest that the answer is a qualified yes, since the residual effects of caste-group belonging are robust to the inclusion of hitherto unobserved credit histories for most, but not all specifications. Provided bank lending is not strongly predicated on good credit histories, the gap between STs and non-STs in most cases remains on average (in the sense of median) in the vicinity of 5% even after accounting for credit histories. This remains so (for $b = 0.4$ and $b = 0.6$) even if the proportion of STs with good credit histories is 0.1 lower than the corresponding proportion for non-STs.

4.2 Small farmers

The Reserve Bank of India has long recognised the challenges faced by small and marginal farmers in accessing agricultural loans. Its policy guidance has encouraged banks to set aside a proportion of overall lending for the 'priority sector', and within this certain 'weaker sections', which include small farmers (Reserve Bank of India, 2007). Small farmers are defined by land holdings of 5 acres or less. Therefore, we next estimate our results separately for small farmers and for farmers who own more than 5 acres of land, to examine whether there is supportive evidence towards this policy's effectiveness.

The top half of table 4 presents these results. As in the preceding section, the model specification used includes all available covariates, and appendix A presents the full regression results. We can see that inter-caste differences in application rates are stark only for large farmers. For small farmers there exist minor differences suggestive of the pattern we witness in the full sample, but their magnitude is far smaller. Similarly, looking at loan approvals, potential discrimination against STs is apparent only for large farmers; for small farmers the loan approval rates are indistinguishable across caste groups.

Figure 1: Simulation results



This suggests that RBI guidance for lending to weaker sections may well be proving effective, even though our data do not allow us to examine this question directly. Not only are loan approval rates essentially equal across caste groups for small farmers, they are also slightly higher than the corresponding rates for large farmers. The same goes for loan application rates, in that small farmers are slightly *more* likely to apply for loans, and significant inter-caste differences exist only for large farmers. It would appear that banks use less discretion in lending to small farmers, and having experienced this willingness over time, small farmers are more likely to apply for loans.

4.3 States with significant ST populations

We now analyse our findings further by focusing on two issues of specific policy relevance: the demographic significance of STs, and the role of land ownership.

First, and as table 1 shows, STs are a substantive proportion in a few states but a very small minority in most others. It is possible that patterns of loan application and approval are different for states which have substantive ST populations, therefore we reestimate our results for these and the remaining states separately. Five states in our sample have an ST population proportion of at least 20%: Jharkhand (50%), Chhattisgarh (42%), Madhya Pradesh (21%), Orissa (21%) and Gujarat (19%).[7]

The bottom half of table 4 summarise the loan application and approval probabilities for these five states and the remaining 13 states. The specification used is the same as model 4 in table 2, which uses all available covariates. Full regression results are presented in appendix A.

These probabilities suggest that the lower loan approval rates for STs in the full sample

[7]Population proportions as a whole, and not just farmers, are Jharkhand (37%), Chhattisgarh (41%), Madhya Pradesh (17%), Orissa (18%) and Gujarat (13%). These are still the five states with the highest proportion of STs.

Table 4: Predicted probabilities for subpopulation estimations

	Loan applications				Loan approvals			
	Brahmin	**OBC**	**SC**	**ST**	**Brahmin**	**OBC**	**SC**	**ST**
Small farmers	0.366	0.430	0.350	0.370	0.939	0.939	0.928	0.925
	(0.042)	(0.034)	(0.044)	(0.051)	(0.022)	(0.018)	(0.028)	(0.033)
p-values	0.090				0.937			
Large farmers	0.333	0.371	0.305	0.261	0.908	0.940	0.937	0.852
	(0.019)	(0.015)	(0.027)	(0.026)	(0.019)	(0.012)	(0.017)	(0.037)
p-values	0.001				0.021			
States with significant ST population	0.332	0.316	0.239	0.257	0.964	0.904	0.899	0.817
	(0.033)	(0.020)	(0.030)	(0.026)	(0.022)	(0.015)	(0.030)	(0.040)
p-values	0.054				0.012			
Remaining states	0.314	0.371	0.298	0.267	0.903	0.937	0.926	0.886
	(0.016)	(0.015)	(0.024)	(0.040)	(0.019)	(0.010)	(0.018)	(0.046)
p-values	0.003				0.283			

Notes:
This table shows sample-average predicted probabilities by caste group for respective subpopulations from Logit models using all available covariates. Thus these are the same as model 4 in table 2. Standard errors are given in parentheses, and p-values shown are for Wald tests for the null that the predicted probabilities are equal across all four caste categories.

arise from states where STs *are* a substantial proportion of the population. States where STs are a very small part of the population do have lower approval rates for ST, but these are only marginally lower than those for non-STs, a trend confirmed by the corresponding p-values as well. It is worth noting that this result is similar to Das et al's (2012) finding that poverty reduction rates are slower amongst STs who reside in states with significant ST populations.

Unfortunately we are unable to consider North-Eastern states in our analysis, where STs are a significant proportion of the population and the majority in some. This limitation is due to the sample size available. There are only 33 farmer households in our data who applied for credit in these states, of which 17 received a loan, thus making it impractical to use regression analysis to model any caste-wise differences. That said, these states are also culturally and economically isolated from other parts of India, have sparser bank networks, and are thus likely to be distinct from the 18 states we are focusing on.

5 Discussion

The Government of India's attempts to deepen financial access by directing banks to lend to SC and ST caste-groups are implicitly based on the assumption that barriers to access lie with lenders. Our analysis partially bears this assumption out, since we find that ST group borrowers are, after controlling for other characteristics, less likely to receive a loan conditional on having applied for one (85% versus 90-93% for the other three caste groups).

However, our analysis also indicates that such directed lending will not alleviate a major component of inter-caste differences in financial access, since there are corresponding, systemic differences between caste groups in loan application rates. Specifically, SCs and STs are less likely to apply for agricultural loans compared to OBC and higher castes. This pattern remains after controlling for various household characteristics, and is qualitatively similar to that of disadvantage and difference in these characteristics. If we take the view that these application rates are at least in part driven by expectations, then it follows that expectations too differ by caste-group in ways similar to incomes, land ownership, or education.

Gary Becker's definition of discrimination places the onus on lenders, wherein discrimination is said to occur if race or other group-based differences remain even after accounting for all characteristics that are relevant to the lending decision. By this definition, if members of one group are characterised by having fewer loans on average because they are less likely to seek credit, these differences would not represent discrimination. Our findings thus indicate that discrimination does likely take place, but while no less serious, it is limited essentially to one caste group (STs). The other inter-caste differences in loan incidence are effectively explained by differences in application rates.

Yet in the Indian context, Dréze et al. (1997) have argued that the distinction between lender discrimination and differences in application rates is largely irrelevant, because any hesitation to apply for a loan in fact reflects differences in expectations that are based on experiences that might span several years or more. That is, lower-caste persons might apply less frequently because they have learned, possibly as a community if not as individuals, that they are less likely to receive loans. In this view caste-wise differences would still reflect discrimination once lending-relevant characteristics are taken into account, but crucially, irrespective of whether a loan was actually applied for or not.

This is an important perspective for understanding the nature and extent of caste-based differences. Nonetheless, we believe that it is crucial to consider loan applications and approvals separately, because the role of caste has different policy implications for each process.

Our finding that caste-wise patterns are largely explained by application rates suggests that policies such as subsidised interest rates or the lowering of collateral requirements are unlikely to reduce all inter-caste differences in credit access. In the short term, loan application rates will be the binding constraint on financial inclusion, and this will improve only once the expectations underlying the demand for credit witness change. The latter changes could take several years whereby these policies, complemented by policies in other realms of economic life, would gradually change borrowers' expectations, helping dissipate the hesitation to seek credit.

However, our finding that STs face lower rates of loan approval suggest that a focus on banks' lending practices, potentially coupled with incentives to lend to ST borrowers, could yield swift improvements for this group. Such policies would be relevant mainly for states where STs are a significant proportion of the population, since we have shown that it is in these states that STs face consistent disadvantage in access to credit. Caste-wise differences in credit access due to banks' unwillingness to lend to STs, potentially due to discrimination, could thus be alleviated. STs would still have lower rates of financial access on account of lower application rates, but overall access would still improve over what is currently witnessed.

Finally, our finding that caste-wise differences are significantly muted for small farmers – those who own less than 5 acres of land – is consistent with the RBI's guidance on lending. We cannot support this claim directly, but it would seem that the RBI's long-standing encouragement of lending to weaker sections of society, of which small farmers are a part, is effective. Not only do we find no evidence of caste discrimination in lending to small farmers, but crucially, it seems that expectations about obtaining loans are also positive, so much so that small farmer SCs and STs are more likely to apply for loans relative to the full sample.

References

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. Volume 3, Part C, pp. 3143–3259. Elsevier.

Asiedu, E., Freeman, J. A. and Nti-Addae, A. (2012). Access to credit by small businesses: How relevant are race, ethnicity, and gender? *American Economic Review 102*(1), 532–537.

Becker, G. (1971). *The economics of discrimination.* London: University of Chicago Press.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010). What's advertising content worth? Evidence from a consumer credit marketing field experiment. *The Quarterly Journal of Economics 125*(1), 263–306.

Blanchard, L., B. Zhao, and J. Yinger (2008, March). Do lenders discriminate against minority and woman entrepreneurs? *Journal of Urban Economics 63*(2), 467–497.

Blanchflower, D. G., P. B. Levine, and D. J. Zimmerman (2003, November). Discrimination in the Small-Business Credit Market. *The Review of Economics and Statistics 85*(4), 930–943.

Chevalier, A. (2007). Education, occupation and career expectations: determinants of the gender pay gap for UK graduates*. *Oxford Bulletin of Economics and Statistics 69*(6), 819–842.

Conning, J. and C. Udry (2007). Rural financial markets in developing countries. In Robert Evenson, Prabhu Pingali, and T. Paul Schultz (Eds.), *Handbook of agricultural economics* (3 ed.)., pp. 2857–2908. Amsterdam: Elsevier.

Desai, S. and R. Vanneman (2015). India Human Development Survey-II (IHDS-II), 2011-12. ICPSR36151-v2.

Deshpande, A. (2000, May). Does Caste Still Define Disparity? A Look at Inequality in Kerala, India. *The American Economic Review 90*(2), 322–325.

Dréze, J., P. Lanjouw, and N. Sharma (1997). Credit in rural India: a case study.

Becker, G. (1971). *Report of the expert group on agricultural indebtedness.* New Delhi: Ministry of Finance.

Ichino, A., F. Mealli, and T. Nannicini (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics 23*(3), 305–327.

Ito, T. (2009, March). Caste discrimination and transaction costs in the labor market: Evidence from rural North India. *Journal of Development Economics 88*(2), 292–300.

Kijima, Y. (2006, January). Caste and Tribe Inequality: Evidence from India, 1983âĂŞ1999. *Economic Development and Cultural Change 54*(2), 369–404.

Kochar, A. (1997). An empirical investigation of rationing constraints in rural credit markets in India. *Journal of Development Economics 53*(2), 339–372.

Kumar, S. M. (2013). Does Access to Formal Agricultural Credit Depend on Caste? *World Development 43*, 315–328.

Maitreyi Bordia Das, Gillette H. Hall, Soumya Kapoor, and Denis Nikitin (2012, April). India. In G. H. Hall and H. A. Patrinos (Eds.), *Indigenous Peoples, Poverty, and Development*, pp. 205–248. Cambridge University Press.

Mishra, S. (2008). Risks, farmers' suicides and agrarian crisis in India? *Indian Journal of Agricultural Economics 63*, 38–54.

National Sample Survey Organisation (2005). Household indebtedness in India as on 30.06.2002. Technical report, National Sample Survey Organisation, Ministry of Statistics and Programme Implementation, Government of India, New Delhi.

Pal, S. (2002). Household Sectoral Choice and Effective Demand for Rural Credit in India. *Applied Economics 34*(14), 1743–1755.

Prakash, A. (2015). *Dalit Capital: State, Markets and Civil Society in Urban India*. New Delhi: Routledge.

Ramaiah, A. (1992, June). Identifying Other Backward Classes. *Economic and Political Weekly 27*(23), 1203–1207.

Reserve Bank of India (2004). *Master circular: Priority sector lending facilities to Scheduled Castes (SCs) & Scheduled Tribes (STs)*. Mumbai: Reserve Bank of India.

Reserve Bank of India (2007). Report of the Working Group to Examine the Procedures and Processes of Agricultural Loans. Technical report, Reserve Bank of India, Mumbai.

Reserve Bank of India (2008). *Master circular - Lending to priority sector*. Mumbai: Reserve Bank of India.

Reserve Bank of India (2014). Priority Sector Lending - Targets and Classification.

Rosenbaum, P. R. and D. B. Rubin (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, 212–218.

Sriram, M. S. (2007, October). Productivity of Rural Credit. *International Journal of Rural Management 3*(2), 245 –268.

Swain, R. B. (2007, December). The demand and supply of credit for households. *Applied Economics 39*(21), 2681–2692.

Thorat, S. and P. Attewell (2007). The Legacy of Social Exclusion: A correspondence study of job discrimination in India. *Economic and Political Weekly*, 4141–4145.

A Logit regression results

Tables 5 and 6 show the logit regression results for the full sample loan applications and approvals, respectively. All models contain state fixed effects and robust standard errors that account for village-level clustering.

Tables 7 and 8 shows subpopulation-wise Logit regression results for loan applications and approvals, respectively. These subpopulations are defined according to land ownership and ST population proportions. These are the full regression results for sections 4.3 and 4.2. All models contain state fixed effects and robust standard errors that account for village-level clustering.

Table 5: Logit regression results: loan applications

	Model 1	Model 2	Model 3	Model 4
Brahmin			(base)	
OBC	0.0209	0.200*	0.231**	0.227**
	(0.105)	(0.103)	(0.104)	(0.104)
SC	-0.589***	-0.221	-0.167	-0.177
	(0.134)	(0.139)	(0.141)	(0.141)
ST	-0.761***	-0.421***	-0.326*	-0.293*
	(0.152)	(0.162)	(0.167)	(0.174)
(log) Land owned		0.583***	0.569***	0.568***
		(0.0564)	(0.0554)	(0.0544)
(log) Land rented in		0.169***	0.175***	0.171***
		(0.0547)	(0.0537)	(0.0542)
(log) Land rented out		-0.138*	-0.143**	-0.144**
		(0.0713)	(0.0719)	(0.0727)
Cons exp		0.457***	0.329***	0.338***
		(0.0684)	(0.0824)	(0.0831)
HH income		-0.000000439*	-0.000000505**	-0.000000504**
		(0.000000225)	(0.000000228)	(0.000000228)
age HH head			-0.000363	-0.000246
			(0.00317)	(0.00317)
sex HH head			0.620***	0.621***
			(0.225)	(0.227)
Years of edu highest male			0.0489***	0.0486***
			(0.00954)	(0.00948)
Years of edu highest female			0.00978	0.00961
			(0.00967)	(0.00956)
HH size			-0.0128	-0.0141
			(0.0193)	(0.0194)
HH prop adult males			-0.269	-0.293
			(0.281)	(0.280)
HH prop adult females			-0.0299	-0.00953
			(0.318)	(0.317)
distance to nearest town				0.134*
				(0.0749)
% HHs with electricity				0.0858*
				(0.0461)
distance closest bank branch				0.0470
				(0.0540)
Constant	-1.925***	-8.014***	-7.406***	-8.399***
	(0.228)	(0.789)	(0.942)	(1.068)
N	7510	7510	7510	7510

Standard errors in parentheses and account for village-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All models contain state fixed effects

Table 6: Logit regression results: loan approvals

	Model 1	Model 2	Model 3	Model 4
Brahmin		(base)		
OBC	0.0940	0.224	0.327	0.327
	(0.277)	(0.266)	(0.259)	(0.264)
SC	-0.194	0.0329	0.262	0.257
	(0.356)	(0.348)	(0.329)	(0.330)
ST	-0.998***	-0.790**	-0.615*	-0.613*
	(0.368)	(0.359)	(0.364)	(0.343)
(log) Land owned		0.537***	0.552***	0.559***
		(0.161)	(0.157)	(0.160)
(log) Land rented in		0.421***	0.448***	0.450***
		(0.146)	(0.144)	(0.144)
(log) Land rented out		-0.158	-0.190	-0.200
		(0.165)	(0.178)	(0.174)
Cons exp		-0.0673	-0.0114	-0.00879
		(0.175)	(0.176)	(0.175)
income		-6.84e-08	-4.61e-08	-5.69e-08
		(0.000000370)	(0.000000392)	(0.000000378)
Age HH head			-0.0174**	-0.0174**
			(0.00754)	(0.00758)
sex HH head			0.403	0.406
			(0.335)	(0.335)
Years of edu highest male			0.00103	0.0000883
			(0.0281)	(0.0283)
Years of edu highest female			0.0540*	0.0538*
			(0.0310)	(0.0311)
HH size			-0.0744*	-0.0725*
			(0.0402)	(0.0405)
HH propn adult males			1.798*	1.800*
			(0.925)	(0.936)
HH propn adult females			-0.158	-0.209
			(0.961)	(0.961)
distance to nearest town				0.0283
				(0.179)
% HHs with electricity				-0.0207
				(0.116)
distance closest bank branch				-0.131
				(0.115)
Constant	2.815***	2.622	1.781	2.030
	(0.785)	(2.062)	(2.216)	(2.444)
N	2587	2587	2587	2587

Standard errors in parentheses and account for village-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All models contain state fixed effects

Table 7: Subpopulation logit regression: loan applications

	Small farmers	Large farmers	Tribal states	Remaining states
Brahmin		(base)		
OBC	0.385*	0.204*	-0.0896	0.319***
	(0.211)	(0.114)	(0.191)	(0.120)
SC	-0.103	-0.157	-0.552**	-0.0946
	(0.234)	(0.180)	(0.263)	(0.164)
ST	0.0219	-0.413**	-0.434*	-0.285
	(0.317)	(0.197)	(0.251)	(0.271)
(log) land owned	1.155***	0.442***	0.599***	0.562***
	(0.201)	(0.0807)	(0.0905)	(0.0661)
(log) land rent in	0.259**	0.170***	-0.0644	0.255***
	(0.108)	(0.0634)	(0.0775)	(0.0676)
(log) land rented out	-0.952**	-0.124*	-0.0837	-0.173*
	(0.459)	(0.0732)	(0.0993)	(0.0945)
cons. exp.	0.144	0.400***	0.616***	0.240**
	(0.141)	(0.105)	(0.126)	(0.1000)
income	4.46e-08	-0.000000487**	-0.00000108***	-0.000000254
	(0.000000332)	(0.000000231)	(0.000000414)	(0.000000211)
age HH head	0.00880	-0.00409	0.000761	-0.00130
	(0.00627)	(0.00366)	(0.00473)	(0.00396)
sex HH head	0.110	0.775***	0.464	0.680**
	(0.425)	(0.254)	(0.306)	(0.281)
Years of edu highest male	0.0545***	0.0468***	0.0547***	0.0482***
	(0.0182)	(0.0116)	(0.0153)	(0.0113)
Years of edu highest female	0.00780	0.00880	-0.0277	0.0206*
	(0.0191)	(0.0116)	(0.0177)	(0.0111)
HH size	0.0104	-0.0221	-0.0420	-0.00800
	(0.0427)	(0.0215)	(0.0292)	(0.0232)
HH propn adult males	-0.0950	-0.425	-0.662	-0.166
	(0.547)	(0.323)	(0.442)	(0.353)
HH propn adult females	-0.263	0.0980	0.507	-0.240
	(0.606)	(0.385)	(0.531)	(0.395)
distance to nearest town/city	0.0437	0.175**	-0.0840	0.237**
	(0.122)	(0.0854)	(0.115)	(0.0951)
% HH with electricity	0.0577	0.106*	0.115	0.0667
	(0.0668)	(0.0602)	(0.0854)	(0.0541)
distance to nearest bank	-0.00965	0.0678	0.00524	0.0614
	(0.0892)	(0.0606)	(0.0934)	(0.0651)
Constant	-6.931***	-8.513***	-10.73***	-7.632***
	(1.748)	(1.361)	(1.486)	(1.288)
N	2489	5021	2459	5051

Standard errors in parentheses and account for village-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All models contain state fixed effects

Table 8: Subpopulation logit regression: loan approvals

	Small farmers	Large farmers	Tribal states	Remaining states
Brahmin		(base)		
OBC	-0.0111	0.518	-1.423*	0.504*
	(0.487)	(0.317)	(0.830)	(0.293)
SC	-0.241	0.450	-1.511*	0.325
	(0.616)	(0.399)	(0.914)	(0.360)
ST	-0.297	-0.598	-2.447***	-0.198
	(0.668)	(0.377)	(0.853)	(0.533)
(log) land owned	1.264***	0.358	1.013***	0.456**
	(0.430)	(0.251)	(0.350)	(0.187)
(log) land rent in	0.899***	0.401**	0.612*	0.434**
	(0.275)	(0.197)	(0.357)	(0.169)
(log) land rented out	0.0361	-0.148	-0.0457	-0.275
	(0.679)	(0.184)	(0.222)	(0.204)
cons. exp.	-0.162	0.122	-0.377	0.0118
	(0.450)	(0.207)	(0.353)	(0.206)
income	0.00000333	-0.000000380	-0.00000202***	0.000000193
	(0.00000256)	(0.000000565)	(0.000000716)	(0.000000733)
age HH head	-0.0134	-0.0192**	-0.0299**	-0.0160*
	(0.0142)	(0.00912)	(0.0144)	(0.00941)
sex HH head	1.193**	0.0794	-0.514	0.659*
	(0.563)	(0.487)	(0.684)	(0.365)
Years of edu highest male	0.00196	0.0195	-0.0681	0.0125
	(0.0422)	(0.0368)	(0.0488)	(0.0325)
Years of edu highest female	0.147***	0.000274	0.0789	0.0465
	(0.0516)	(0.0366)	(0.0502)	(0.0369)
HH size	-0.278***	-0.00664	0.117	-0.0929**
	(0.0875)	(0.0539)	(0.101)	(0.0444)
HH propn adult males	2.766*	0.955	1.778	1.915
	(1.511)	(1.178)	(1.437)	(1.199)
HH propn adult females	-0.0335	0.0760	1.398	-0.511
	(1.768)	(1.140)	(1.529)	(1.180)
Distance nearest town/city	-0.429	0.269	-0.152	0.148
	(0.318)	(0.208)	(0.305)	(0.215)
% HH with electricity	-0.343**	0.148	0.0951	-0.120
	(0.175)	(0.146)	(0.202)	(0.105)
Distance nearest bank	-0.125	-0.161	-0.251	-0.109
	(0.199)	(0.153)	(0.205)	(0.139)
Constant	3.993	15.20	3.282	1.738
	(5.780)	(.)	(4.404)	(2.864)
N	476	2111	854	1733

Standard errors in parentheses and account for village clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions contain state fixed effects.

B Details of the simulation exercise

Figure 2 summarises the values of $\{b, \mu_{ST}, ST_{gap}\}$ that satisfy the following constraints:

$$1 \geq p_{01} \geq 0 \quad \Leftrightarrow \quad 1 \geq \frac{\mu_{ST} - b\theta_{11}}{\theta_{01}} \geq 0 \tag{12}$$

$$1 \geq p_{00} \geq 0 \quad \Leftrightarrow \quad 1 \geq \frac{\mu_{ST} + ST_{gap} - b\theta_{10}}{\theta_{00}} \geq 0 \tag{13}$$

$$1 \geq b \geq 0 \tag{14}$$

$$1 \geq \mu_{ST} \geq 0 \tag{15}$$

$$1 \geq \mu_{ST} + ST_{gap} \geq 0 \tag{16}$$

Figure 2: Feasible parameter values for the simulation exercise

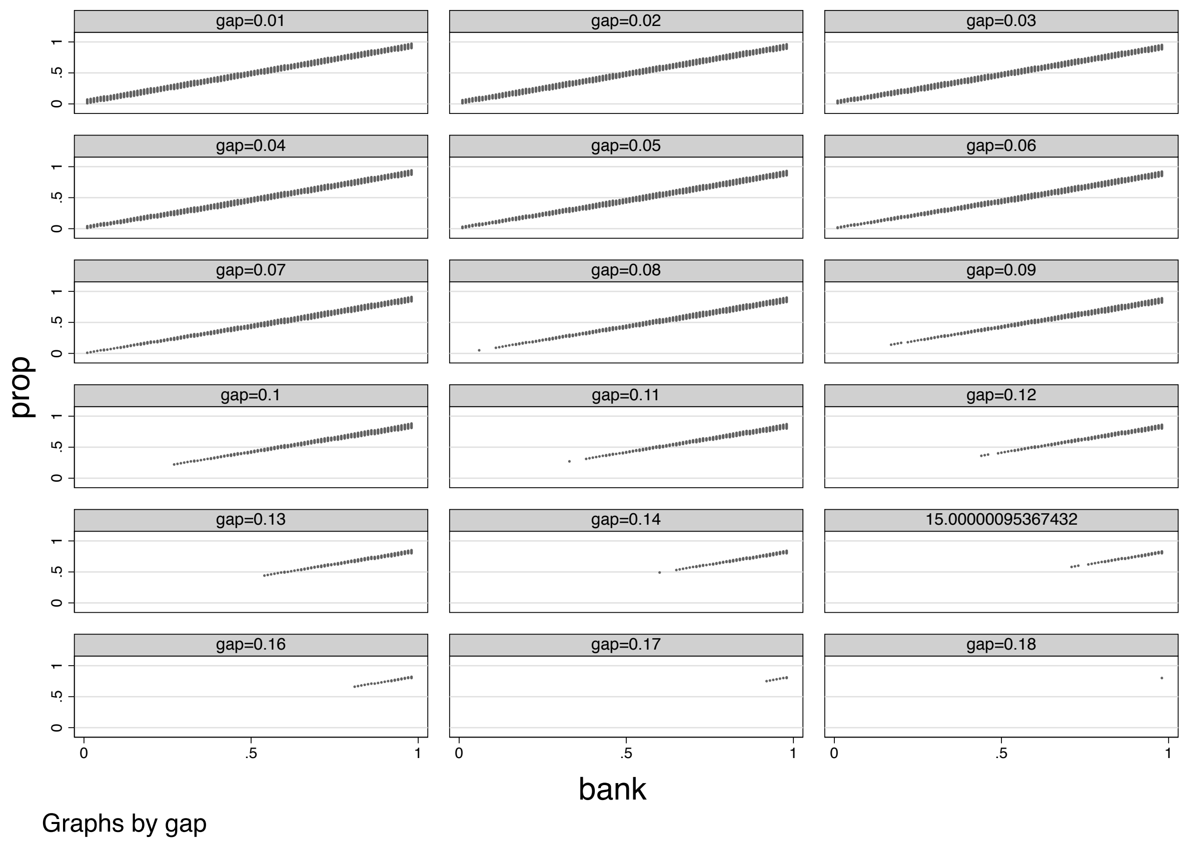


Figure 3 is similar to figure 1 in the main text, and presents a slightly different way of summarising the results from our simulation exercise. Instead of calculating the difference in predicted probabilities of loan approval for STs and the average of non-STs (Brahmins, OBCs, SCs), here we calculate the difference between average predicted probabilities for STs and the minimum of those for Brahmins, OBCs and SCs. This difference is thus a more conservative estimate of the ST vs non-ST gap in loan approvals. The distribution of these differences across 1000 simulations for each combination of $\{b, ST_{gap}\}$ are summarised using box plots.

Figure 3: Simulation results